International Journal for Parasitology 44 (2014) 625–635

Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara

Ligand heterogeneity of the cysteine protease binding protein family in the parasitic protist *Entamoeba histolytica*

Konomi Marumo [a,b,1], Kumiko Nakada-Tsukui [a,1], Kentaro Tomii [c], Tomoyoshi Nozaki [a,b,*]

[a] Department of Parasitology, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan
[b] Graduate School of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan
[c] Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

ARTICLE INFO

Article history:
Received 27 February 2014
Received in revised form 11 April 2014
Accepted 15 April 2014
Available online 5 June 2014

Keywords:
Amylase
β-Hexosaminidase
Cysteine protease
Entamoeba histolytica
Lysosomes
Receptor

ABSTRACT

Lysosomal soluble proteins are targeted to endosomes and lysosomes by specific receptors resident in the endoplasmic reticulum and/or the Golgi apparatus. The enteric protozoan parasite *Entamoeba histolytica* has a novel class of lysosomal targeting receptors, named the cysteine protease binding protein family (CPBF). Among 11 CPBFs (CPBF1–11), ligands for three members, CPBF1, CPBF6 and CPBF8, were previously shown to be cysteine proteases, α- and γ- amylases, and β-hexosaminidase and lysozymes, respectively. To further understand the heterogeneity of the ligands of CPBFs, we attempted to isolate and identify the ligands for other members of CPBFs, namely CPBF2, 3, 4, 5, 7, 9, 10 and 11, by immunoprecipitation and mass spectrometric analysis. We found that CPBF2 and CPBF10 bound to α-amylases while CPBF7 bound to β-hexosaminidases. It is intriguing that cysteine protease are exclusively recognised by CPBF1, whereas three α-amylases and β-hexosaminidases are redundantly recognised by three and two CPBFs, respectively. It was shown by bioinformatics analysis and phylogenetic reconstruction that each CPBF contains six prepeptidase carboxyl-terminal domains, and the domain configuration is evolutionarily conserved among CPBFs. Taken together, CPBFs with unique and conserved domain organisation have a remarkable ligand heterogeneity toward cysteine protease and carbohydrate degradation enzymes. Further structural studies are needed to elucidate the structural basis of the ligand specificity.

1. Introduction

Lysosomal enzymes such as cysteine proteases (CPs) play a pivotal role in the pathogenesis of the intestinal parasitic protist *Entamoeba histolytica*. Cytolytic capacity and tissue invasiveness of this parasite are mainly attributed to CPs, as shown in numerous in vitro and in vivo studies (Brinen et al., 2000; Que and Reed, 2000; Hellberg et al., 2001, 2002; Bruchhaus et al., 2003; Que et al., 2003; Ackers and Mirelman, 2006; Gilchrist et al., 2006; MacFarlane and Singh, 2006; Meléndez-López et al., 2007; He et al., 2010; Ralston and Petri, 2011). The regulation of their intracellular processing and transport has begun to be unveiled by our recent discovery of the novel CP-specific carrier/receptor protein, named cysteine protease binding protein family (CPBF) 1 (Nakada-Tsukui et al., 2012). CPBF1 is a unique cargo receptor restricted to the Amoebozoa, and shows a number of differences from known transport receptors in other eukaryotic lineages.

In general, transport of soluble lysosomal proteins is mediated by three major classes of soluble lysosomal protein transport receptors: mannose 6-phosphate receptor (MPR), sortilin or vacuolar protein sorting 10 protein (Vps10p), and plant-specific vacuolar sorting receptor (VSR). Sortilin/Vps10p is conserved in a wide range of eukaryotes, while MPR is mainly conserved among the Opisthokonta and VSR is specific to the Planta and the Chloroplastida. MPRs consist of two classes of proteins, cation-independent MPR (CI-MPR) and cation-dependent MPR (CD-MPR), and recognise the mannose 6-phosphate moiety on the soluble lysosomal proteins via its carbohydrate recognition domain (CRD). There are two genes encoding putative CD-MPR in *E. histolytica*. However, immunoprecipitation of influenza virus hemagglutinin (HA)-tagged CD-MPRs demonstrated no interaction with soluble lysosomal proteins (Nakada-Tsukui et al., unpublished data), suggesting that MPRs are unlikely to function as lysosomal targeting receptors in *E. histolytica*. Furthermore, neither Sortilin/Vps10p nor VSR is present in the genome.

* Corresponding author at: Department of Parasitology, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan. Tel.: +81 3 4580 2690; fax: +81 3 5285 1173.
E-mail address: nozaki@nih.go.jp (T. Nozaki).
[1] These two authors equally contributed to this work.

http://dx.doi.org/10.1016/j.ijpara.2014.04.008
0020-7519/

In *E. histolytica*, CPBF consists of 11 members with 18–75% mutual amino acid identities. We previously demonstrated that three most highly expressed CPBFs, CPBF1, CPBF6, and CPBF8, are involved in the targeting of soluble lysosomal proteins including CP, amylases, β-hexosaminidase and lysozymes (Furukawa et al., 2012, 2013; Nakada-Tsukui et al., 2012). As MPR, Sortilin/Vps10p and VSR are generally encoded by a single gene in the genome, CPBF represents the first protein family involved in targeting of lysosomal enzymes. All members of CPBFs share similar features such as the signal sequence at the amino terminus, a single transmembrane domain and the YxxΦ motif at the carboxyl terminus. The YxxΦ motif (x is any amino acid and Φ is any aliphatic amino acid) is known to be present in the cytoplasmic portion of numerous receptors and responsible for binding to the adaptor protein (AP) complex (Nakatsu and Ohno, 2003). These common features suggest that all members of CPBF are involved in lysosomal targeting of respective specific soluble lysosomal proteins. To further examine the specificity and heterogeneity of the ligands of other members of CPBFs, we attempted to identify and characterise the ligands for CPBF2, 3, 4, 5, 7, 9, 10 and 11 by immunoprecipitation and mass spectrometric analysis.

2. Materials and methods

2.1. Cells and reagents

Trophozoites of *E. histolytica* strain HM-1:IMSS cl6 (HM-1) were cultured axenically in BI-S-33 medium (Diamond et al., 1978) at 35.5 °C, as previously described (Clark and Diamond, 2002). Amoeba transformants were cultured in the presence of 10 μg/mL of Geneticin. *Escherichia coli* strain DH5α was purchased from Life Technologies (Tokyo, Japan). All chemicals of analytical grade were purchased from Sigma–Aldrich (Tokyo, Japan) unless otherwise stated.

2.2. Plasmid construction

Standard techniques were used for routine DNA manipulation, subcloning and plasmid construction (Sambrook and Russell, 2001). Plasmids to express CPBF2, 3, 4, 5, 7, 9, 10 or 11 fused with the HA epitope at the carboxyl terminus were generated by the insertion of the corresponding protein coding region of the *CPBF* gene into the *Bgl*II site of a pEhExHA vector (Nakada-Tsukui et al., 2009) either by standard restriction digestion and ligation methods for CPBF3, 4, 10 and 11, or by InFusion system (Takara, Tokyo, Japan) for CPBF2, 5, 7 and 9. Resultant plasmids were named pEhExHA-CPBF2, 3, 4, 5, 7, 9, 10 and 11, respectively. The protein coding region of each *CPBF* gene was amplified with specific sense and antisense oligonucleotide primers: acacattaacAGATCATGGTTGTTCTGTTTTTATT and atggatacatAGATCGAAAGTTCCAAATGATGATT (CPBF2); accggatccATGATCCTATTAATTCTAGCA and gttggatccAAGTTCATGATATCCCAAAAA (CPBF3); accggatccATGGTCCAAATAACATGTCTT and gttggatccAAGTTCATGATATCTCAATAA (CPBF4); acacattaacAGATCATGTTTATTCTTCTTAGTCT and atggatacatAGATCAAAGTCAGAATAACTCTTTC (CPBF5); acacattaacAGATCATGTTGGTTTTCTTAACAAT and atggatacatAGATCAACTAAAGTAGCATATCCAG (CPBF7); acacattaacAGATCATGTTATTGAAATGGGGATT and atggatacatAGATCATTATCAATAATTGTTTTTA (CPBF9); accggatccATGCTTTTAATAACTCTCCTC and gttggatccGAAACTACTGAAACTTGATGA (CPBF10); accggatccATGTTTTTGTTGTTCATTTCT and gttggatccTAATTCATAATATCCTTTGTT (CPBF11). Plasmids to express GST-fusion proteins with the individual prepeptidase carboxyl-terminal (PPC) domain (PPC1-6) of CPBF1 were generated by the insertion of the synthesized nucleotides corresponding to CPBF1 PPC1-6 or the first PPC domain of CPBF8 (CPBF8 PPC1) into the *Bam*HI and *Not*I double-digested pGEX6p-2 vector (GE Healthcare, Tokyo, Japan), and designated as pGST-CPBF1 PPC1-6 or pGST-CPBF8 PPC1, respectively. CPBF1 PPC1-6 corresponds with amino acids (a.a.) 20∼165, 172∼298, 303∼428, 435∼570, 574∼710, and 717∼853, of CPBF1, respectively, and CPBF8 PPC1 corresponds to a.a. 16∼154 of CPBF8.

2.3. Amoeba transformation

pEhExHA-CPBF2, 3, 4, 5, 7, 9, 10 or 11 was introduced into HM-1 trophozoites by lipofection, as previously described (Nozaki et al., 1999). Geneticin was added at a concentration of 1 μg/mL at 24 h after transfection and gradually increased for approximately 2 weeks until the G418 concentration reached 10 μg/mL.

2.4. Immunoprecipitation, SDS–PAGE and immunoblot analyses

For the isolation of CPBF-HA binding proteins, the cell pellet from 2.0×10^7 CPBF-HA-expressing or mock-transfected cells was lysed with 1 mL of lysis buffer (50 mM Tris–HCl pH. 7.5, 150 mM NaCl, 1% Triton-X100, 0.5 mg/mL of E-64, complete mini EDTA-free protease inhibitor cocktail (Roche Applied Science, Penzberg, Germany)). After centrifugation at 14,000*g* for 5 min at 4 °C, the soluble lysate was pre-cleared with 50 μL of protein G Sepharose (50% slurry in lysis buffer), (GE Health Care, Waukesha, WI, USA) and then mixed and incubated with 50 μL of anti-HA monoclonal antibody-conjugated agarose (Sigma–Aldrich, St. Louis, MO, USA) for 3.5 h at 4 °C. Immune complexes bound to the resin were washed five times with wash buffer (50 mM Tris–HCl pH. 7.5, 150 mM NaCl, 1% Triton-X100) and then eluted by incubating the resin with 180 μL of 200 mg/mL HA peptide (Sigma–Aldrich) in lysis buffer for 16 h at 4 °C. Approximately 2 μg of the eluted samples were subjected to SDS–PAGE and visualised with either a silver stain MS kit (WAKO, Tokyo, Japan) or a SYPRO ruby protein stain (Takara). The same samples were also subjected to SDS–PAGE and immunoblot analyses as previously described (Sambrook and Russell, 2001). Primary antibodies were used at a 1:500 dilution for anti-Cm-EhCP-A5 rabbit antibody (Nakada-Tsukui et al., 2012) or at a 1:1000 dilution for anti-HA mouse monoclonal antibody (clone 11MO, Covance, Princeton, NJ, USA) in immunoblot analyses. CP-A5 is the major CP that CPBF1 was found to bind (Nakada-Tsukui et al., 2012).

2.5. Mass spectrometric analysis

Unique bands detected exclusively in the eluted samples from the HA-tagged transformants but not those from the control, after visualisation by silver or SYPRO ruby stain, were excised and subjected to LC-MS/MS analysis. The total mixture of the immunoprecipitated eluates using the lysate from CPBF2, 3, 4, 5, 7, 9, 10 and 11-HA expressing and mock transformants were briefly electrophoresed on SDS–PAGE to allow entry of proteins into the gel, visualised by silver stain, and the bands containing whole mixture were excised and subjected to LC–MS/MS analysis.

LC–MS/MS analysis was performed at W. M. Keck Biomedical Mass Spectrometry Laboratory, University of Virginia, USA. The gel pieces from the band were transferred to a siliconized tube and washed in 200 μL of 50% methanol. The gel pieces were dehydrated in acetonitrile, rehydrated in 30 μL of 10 mM DTT in 0.1 M ammonium bicarbonate and reduced at room temperature for 0.5 h. The DTT solution was removed and the sample alkylated in 30 μL of 50 mM iodoacetamide in 0.1 M ammonium bicarbonate at room temperature for 0.5 h. The reagent was removed and the gel pieces dehydrated in 100 μL of acetonitrile. The acetonitrile was removed and the gel pieces rehydrated in 100 μL of 0.1 M ammonium bicarbonate. The pieces were dehydrated in 100 μL

of acetonitrile, the acetonitrile removed and the pieces completely dried by vacuum centrifugation. The gel pieces were rehydrated in 20 ng/μL of trypsin in 50 mM ammonium bicarbonate on ice for 30 min. Any excess enzyme solution was removed and 20 μL of 50 mM ammonium bicarbonate added. The sample was digested overnight at 37 °C and the peptides formed extracted from the polyacrylamide in a 100 μL aliquot of 50% acetonitrile/5% formic acid. This extract was evaporated to 15 μL for MS analysis. The LC–MS system consisted of a Thermo Electron Velos Orbitrap ETD mass spectrometer system with a Protana nanospray ion source interfaced to a self-packed 8 cm x 75 μm inner diameter Phenomenex Jupiter 10 μm C18 reversed-phase capillary column. The extract (7 μL) was injected and the peptides eluted from the column by an acetonitrile/0.1 M acetic acid gradient at a flow rate of 0.5 μL/min over 1.2 h. The nanospray ion source was operated at 2.5 kV. The digest was analysed using the double play capability of the instrument, acquiring a full scan mass spectrum to determine peptide molecular weights followed by product ion spectra to determine a.a. sequence in sequential scans.

2.6. Data analysis to determine specific binding proteins

The data were analysed by database searching using the Sequest search algorithm against the *E. histolytica* genome database (http://amoebadb.org/amoeba/). The quantitative value (QV), normalised with unweighted spectrum counts, was used to estimate relative quantities of proteins in the samples. Specific binding proteins were determined by the following criteria. First, proteins that showed QV > 8, or QV > 10 in the control pEhExHA transformed sample ("HA" in Table 1) and proteins that showed QV < 3 in the CPBF samples were removed, and it was assumed that those were non-specific proteins. The proteins that showed >3 or >4-fold higher QV in the CPBF samples compared with those in the HA control were selected. Finally, proteins lacking the signal sequence were removed from a list of possible ligands. Applying these criteria to the proteins discovered, positive controls, i.e., CPs in CPBF1-HA, were unequivocally detected.

2.7. Indirect immunofluorescence assay

The indirect immunofluorescence assay was performed as previously described (Nakada-Tsukui et al., 2012). Briefly, the amoeba transformant cells were harvested and transferred to 8 mm round wells on a slide glass, and then fixed with 3.7% paraformaldehyde in PBS, pH 7.2, for 10 min. After washing, the cells were permeabilized with 0.2% saponin in PBS containing 1% BSA for 10 min, and reacted with an anti-HA monoclonal antibody (clone 11MO, Covance) diluted at 1:1000 in PBS containing 0.2% saponin and 1% BSA. After washing three times with PBS containing 0.1% BSA, the samples were then reacted with Alexa Fluor 488-conjugated anti-mouse secondary antibody (1:1000 dilution in PBS containing 0.2% saponin and 1% BSA) for 1 h. For lysosomal staining, 10 μM Lysotracker Red (Molecular Probes, Eugene, OR, USA) was added to *E. histolytica* transformants for 16 h, and the trophozoites were then washed, harvested and subjected to an immunofluorescence assay. The samples were examined on a Carl-Zeiss LSM 510 META confocal laser-scanning microscope. The resultant images were further analysed using LSM510 software.

2.8. In silico identification of PPC domains in CPBF

To identify structural/functional domains in CPBFs, we utilised our profile-profile alignment methods, called FORTE (Tomii and Akiyama, 2004). FORTE utilises position-specific score matrices (PSSMs) for both the query and library proteins to perform profile-profile alignment. Previous applications in the Critical Assessment of Protein Structure Prediction (CASP) experiments (http://predictioncenter.org/) should also be referred to (Shiozawa et al., 2004; Tomii et al., 2005, 2012; Wang et al., 2005). We created and evaluated a phylogenetic tree of a total of 66 individual domains (six PPC domains in each CPBFs). A multiple alignment of those a.a. sequences was constructed by Clustal Omega (Sievers et al., 2004). The phylogenetic trees were constructed using the neighbour joining method using Clustal W (Larkin et al., 2007). The tree was depicted with NJplot (Perrière and Gouy, 1996).

2.9. Recombinant protein expression and in vitro binding assay

GST-fused recombinant proteins containing individual PPC domains (CPBF1 PPC1-6 and CPBF8 PPC1) were produced as follows: pGST-CPBF1PPC1-6 and pGST-CPBF8PPC1 were introduced into *E. coli* BL21(DE3) competent cells (Merck, Tokyo, Japan). Expression of the recombinant proteins was induced with 100 mM isopropyl-β-thiogalactoside (IPTG) at 25 °C for 5 h. The bacterial cells were collected and lysed by adding bacterial protein extraction reagent in phosphate buffer (B-PER) (Thermo Scientific, Tokyo, Japan) to the cell pellet. Clear lysate was mixed with glutathione Sepharose 4B (GE Healthcare) for 1 h at 4 °C then washed three times with wash buffer (50 mM Tris–HCl pH. 7.5, 150 mM NaCl, 1% Triton-X100). The GST-CPBF PPC-bound Sepharose beads were mixed with the soluble supernatant of lysates prepared from 3×10^6 HM-1 trophozoites as described in Section 2.4 and incubated for 1 h at 4 °C. The beads were washed three times with wash buffer and boiled with SDS–PAGE loading buffer. The eluted proteins were separated by SDS–PAGE and analysed by Coomassie Brilliant Blue stain (CBB, one step CBB stain kit, Bio Craft, Tokyo, Japan) and an immunoblot assay.

Images of CBB-stained polyacrylamide gel and immunoblots were acquired by GELSCAN (iMeasure Inc., Nagano, Japan) and LAS3000 (GE Healthcare), respectively. The O.D. of the bands was quantified using Image J (http://rsbweb.nih.gov/ij/index.html). Binding efficiency was estimated with the parameter defined as (the O.D. of the band corresponding to CP-A5 on an immunoblot) divided by (the O.D. of the GST-fusion protein band on a CBB-stained gel). Relative binding efficiency of each GST-PPC domain fusion protein to CP-A5 was expressed after normalisation against the value of the GST control.

3. Results and discussion

3.1. Establishment of CPBF-HA expressing transformants and potential post-translational modifications of CPBFs

While the ligands of CPBF1, 6 and 8 were identified in our previous studies (Furukawa et al., 2012, 2013; Nakada-Tsukui et al., 2012), the spectrum of the ligands recognised by other members of CPBFs remained poorly understood. Thus, we established *E. histolytica* transformants expressing CPBF2, 3, 4, 5, 7, 9, 10 or 11, tagged with the carboxyl-terminal HA epitope, to identify the ligands of all members of CPBFs. In all experiments, the ameba transformants transfected with a pEhExHA mock vector and a pCPBF1-HA vector were used as negative and positive controls. Expression of HA-fused CPBFs was confirmed by immunoblot analysis with anti-HA antibody (Supplementary Fig. S1).

All of the HA-tagged CPBF proteins showed molecular masses slightly higher than those predicted, as seen for other HA-tagged proteins (Nakada-Tsukui et al., 2005, 2012; Furukawa et al., 2012, 2013). Even if considering the effect of the HA tag, CPBF7 and CPBF10 showed higher molecular masses than other CPBFs, suggesting possible post-translational modifications, similar to

Table 1
Ligands and associated proteins of cysteine protease binding protein family (CPBF) 2–11 identified by immunoprecipitation and LC–MS/MS analysis.

			Accession number		Quantitative value[d]		Unique peptides[e]	
CPBF	Identified proteins	MW	GenBank	AmoebaDB	CPBF	HA	CPBF	HA
					CPBF2		CPBF2	
CPBF2	CPBF2	97 kDa	XP_653276	EHI_087660	347.17	0	43	0
	α-Amylase family protein	69 kDa	XP_655699	EHI_152880	122.46	0	25	0
	70 kDa heat shock protein	73 kDa	XP_654737	EHI_199590	3.37	0	3	0
	Hypothetical protein	24 kDa	XP_655760	EHI_155310	3.37	0	1	0
					CPBF3		CPBF3	
CPBF3	CPBF3	96 kDa	XP_649180	EHI_161650	203.48	0	42	0
	70 kDa Heat shock protein	73 kDa	XP_654737	EHI_199590	14.80	2.64	12	3
	CPBF4	98 kDa	XP_655897	EHI_012340	12.02	0	1	0
					CPBF4		CPBF4	
CPBF4	CPBF4	98 kDa	XP_655897	EHI_012340	152.30	0	34	0
	CPBF3	96 kDa	XP_649180	EHI_161650	15.49	0	2	0
	Serine-threonine-isoleucine rich protein	260 kDa	XP_001913596	EHI_004340	4.30	0	4	0
	EhCP-A2	35 kDa	XP_650642	EHI_033710	3.44	0	3	0
	Galactose-specific lectin light subunit	34 kDa	XP_001913429	EHI_049690	3.44	0	4[f]	0
					CPBF5		CPBF5	
CPBF5	CPBF5	96 kDa	XP_654065	EHI_137940	172.19	0	33	0
	70 kDa heat shock protein	73 kDa	XP_654737	EHI_199590	12.98	0	7	0
	Galactose-specific lectin light subunit	34 kDa	XP_656145	EHI_035690	6.92	0	5	0
	Hypothetical protein	34 kDa	XP_650601	EHI_047800	3.46	0	3	0
CPBF6[a]	CPBF6	99 kDa	XP_653036	EHI_178470				
	α-Amylase family protein	57 kDa	XP_655636	EHI_023360				
	γ-Amylase	75 kDa	XP_652381	EHI_044370				
					CPBF7		CPBF7	
CPBF7	CPBF7	100 kDa	XP_649361	EHI_040440	344.40	3.14	33	0
	β-N-acetylhexosaminidase	64 kDa	XP_656208	EHI_012010	17.95	0	5	0
	β-N-acetylhexosaminidase, subunit	64 kDa	XP_650273	EHI_007330	16.32	0	5	0
	MPR1	24 kDa	XP_656907	EHI_096320	13.06	0	4	0
	Pore-forming peptide ameobapore B precursor	10 kDa	XP_001913632	EHI_194540	9.79	0	3	0
	70 kDa Heat shock protein	73 kDa	XP_654737	EHI_199590	8.16	0	4	0
	Hypothetical protein	30 kDa	XP_652382	EHI_044360	6.53	3.14	2	1
	Hypothetical protein	17 kDa	XP_650886	EHI_069510	3.26	1.57	1	1
	Hypothetical protein	24 kDa	XP_655760	EHI_155310	3.26	0	1	0
	Hypothetical protein	59 kDa	XP_656261	EHI_178650	3.26	0	2	0
CPBF8[b]	CPBF8	100 kDa	XP_652899	EHI_059830				
	β-hexosaminidase, "alpha" sign-subunit	60 kDa	XP_657529/AJ582954[c]	EHI_148130				
	Lysozyme1	23 kDa	XP_653294	EHI_199110				
	Lysozyme2	23 kDa	XP_656933	EHI_096570				
					CPBF9		CPBF9	
CPBF9	CPBF9	100 kDa	XP_655360	EHI_021220	100.27	0	18	0
	Hypothetical protein	18 kDa	XP_656071	EHI_117850	10.29	0	1	0
	70 kDa Heat shock protein	73 kDa	XP_654737	EHI_199590	10.29	0	3	0
	Lysozyme2	23 kDa	XP_656933	EHI_096570	7.71	1.52	1	1
					CPBF10		CPBF10	
CPBF10	CPBF10	98 kDa	XP_649015	EHI_191730	63.88	0	12	0
	α-Amylase	53 kDa	XP_656406	EHI_153100	49.05	0	10	0
	α-Amylase family protein	57 kDa	XP_655636	EHI_023360	27.38	5.76	10	4
	70 kDa Heat shock protein	73 kDa	XP_654737	EHI_199590	25.09	4.61	11	3
	Hypothetical protein	59 kDa	XP_656261	EHI_178650	19.39	0	6	0
	β-Amylase	47 kDa	XP_653896	EHI_192590	17.11	2.31	6	2
	Hypothetical protein	71 kDa	XP_651525	EHI_022130	4.56	0	3	0
	Hypothetical protein	57 kDa	XP_648234	EHI_025100	4.56	0	3	0
	MPR1	24 kDa	XP_656907	EHI_096320	3.42	0	3	0
					CPBF11		CPBF11	
CPBF11	CPBF11	86 kDa	XP_656044	EHI_118120	89.53	0	23	0
	70 kDa Heat shock protein	73 kDa	XP_654737	EHI_199590	24.11	4.61	17	3

HA, hemagglutinin; EhCP, *Entamoeba histolytica* cysteine protease; MPR, mannose 6-phosphate receptor.
[a] From Furukawa et al. (2013).
[b] From Furukawa et al. (2012).
[c] XP_657529 (EHI_148130) and AJ582954 are identical except that XP_657529 (EHI_148130) starts at the second methionine of AJ582954 and lacks the signal sequence.
[d] Quantitative values are shown for the identified proteins from the CPBF-HA and control transformants.
[e] The number of unique peptides detected are shown.
[f] This protein is similar to two other closely related proteins and the number of all detected peptides is shown.

CPBF6 and CPBF8 which have a serine-rich region (SRR) upstream of the transmembrane domain (Furukawa et al., 2012, 2013). It was previously demonstrated that a deletion of the SRR in CPBF8 caused a mobility shift in the predicted molecular masses and a decrease in the ligand binding (Furukawa et al., 2012). In addition, CPBF7 and CPBF10 showed close kinship with CPBF6 and CPBF8 by phylogenetic analysis (Nakada-Tsukui et al., 2012). While CPBF7 has a SRR (Furukawa et al., 2012), there is no apparent SRR in CPBF10; CPBF10 contains only two serine residues within the luminal portion near the transmembrane domain. There is no potential N-glycosylation site, either, as predicted by NetNglyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/).

3.2. Immunoprecipitation of CPBF-binding proteins

All CPBF-HA- and mock-transfected *E. histolytica* lines were subjected to immunoprecipitation with anti-HA antibody, separated by SDS–PAGE and visualised with silver or SYPRO ruby stain (Fig. 1). Immunoprecipitation of the CPBF-HA proteins was confirmed in all transformants. Compared with the mock transfected line ("HA" in Fig. 1), one extra band at approximately 70 kDa in CPBF2-HA, three extra bands at approximately 60, 55 and 40 kDa in CPBF10-HA, and one extra band around 45 kDa in CPBF11-HA were detected (Fig. 1). These specific bands were excised and subjected to LC-MS/MS analysis. We also analysed whole immunoprecipitated samples from lysates of CPBF1, 2, 3, 4, 5, 7, 9, 10, 11 and the mock control by LC–MS/MS. In the following sections, we categorised CPBF members based on their ligand specificities.

3.3. CPBF2, CPBF6 and CPBF10 bound to amylases

Three CPBFs, namely CPBF2 and CPBF10, as well as previously identified CPBF6 (Furukawa et al., 2013), bound to a variety of amylases. Silver staining of immunoprecipitated samples from CPBF2-HA lysates after SDS–PAGE showed a specific 70 kDa band (Fig. 1). LC–MS/MS analysis of the band (Supplementary Table S1) and the whole immunoprecipitated sample (Table 1) indicated the protein

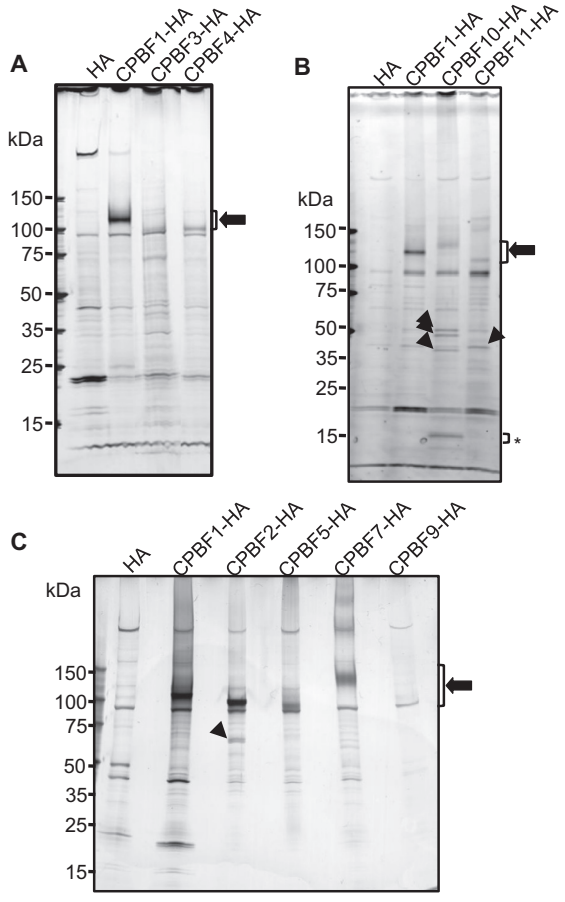

Fig. 1. SDS–PAGE analysis of immunoprecipitated mixtures of *Entamoeba histolytica* cysteine protease binding protein families (CPBFs) and ligands. CPBF1, 2, 3, 4, 5, 7, 9, 10 and 11-haemagglutinin (HA) were immunoprecipitated from the corresponding transformant lines with anti-HA monoclonal antibody, separated by SDS–PAGE and stained by (A, C) silver staining or (B) Sypro Ruby staining. (A) CPBF1, 3 and 4-HA; (B) CPBF1, 10 and 11-HA; (C) CPBF1, 2, 5, 7 and 9-HA. Arrows indicate the bait (CPBF–HA) immunoprecipitated, and arrowheads depict candidates for co-immunoprecipitated ligands. Note that immunoprecipitation and electrophoresis were conducted in three independent experiments. * These bands were not reproducible.

to be α-amylase (XP_655699, EHI_152880), with 22% and 42% coverage, respectively, and a high QV (122.5 for the whole sample).

The 60 and 55 kDa bands exclusively detected in CPBF10-HA were identified as α-amylases (XP_655636 (EHI_023360) and XP_656406 (EHI_153100)), with 23% and 25% coverage, respectively (Supplementary Table S2). These two amylases were also detected in the whole immunoprecipitated sample from CPBF10-HA (Table 1). One should note that these two α-amylases were different from α-amylases that bind to CPBF2 (XP_655699, EHI_152880). The 40 kDa band detected in the immunoprecipitated sample from CPBF10-HA was not unequivocally assigned (QV < 4). Another α-amylase, XP_656406 (EHI_153100), was detected from the 40 kDa band, despite a low QV (3) and being more frequently detected in the 55 kDa band. Intriguingly, one α-amylase (XP_655636, EHI_023360) was also identified as the cargo of CPBF6 (Furukawa et al., 2013), which shows phylogenetic kinship with CPBF10 (Nakada-Tsukui et al., 2012). In addition to these α-amylases, β-amylase, XP_653896 (EHI_192590), was detected from whole mixture.

Three α-amylases found as CPBF ligands in this study were previously detected in our phagosome proteome studies (Okada et al., 2006; Furukawa et al., 2013). A recent transcriptomic analysis using the ex vivo human colon explant showed that trophozoites of the virulent strain showed a remarkable up-regulation of genes implicated in carbohydrate metabolism and processing of glycosylated residues compared with the non-virulent strain (Thibeaux et al., 2013). It was shown in that study that among the carbohydrate metabolism-related genes, β-amylase (XP_653896, EHI_192590) was the most highly induced (approximately 17-fold increase) in the virulent strain compared with the non-virulent strain. Furthermore, Thibeaux et al. (2013) showed that the gene repression of β-amylase caused a reduction in mucus layer degradation. Together with our previous observation of β-amylase localization in phagosomes (Furukawa et al., 2013), these findings suggest a role for amylases and their corresponding CPBF receptors in pathogenesis.

3.4. Polymorphism of amylases

There are at least five independent (non-allelic) α-amylase genes (XP_656406, EHI_153100; XP_655636, EHI_023360; XP_655699, EHI_152880; XP_649162, EHI_130690; XP_652044, EHI_055650). Among these five α-amylases, CPBF2, 6 and 10 bind to three of them (XP_656406, EHI_153100; XP_655636, EHI_023360; and XP_655699, EHI_152880), all of which possess the signal peptide. Among α-amylases that interact with CPBFs, XP_655699 (EHI_152880) and XP_656406 (EHI_153100) specifically interact with CPBF2 and CPBF10, respectively, whereas XP_655636 (EHI_023360) interacts with both CPBF6 and CPBF10. XP_655636 (EHI_023360) is the most highly expressed mRNA among all putative α-amylase genes, as demonstrated by our previous microarray analysis (Penuliar et al., 2012). This is one of the two examples in which one ligand is recognised by more than one CPBF (see below). Although it was previously shown that SRR is essential for the binding of CPBF6 to α- and γ-amylases (Furukawa et al., 2013), CPBF10 appears to lack SRR. Possible post-translational modifications on CPBF10, as suggested by slower migration on SDS–PAGE (see Section 3.1), and their involvement in the ligand interaction needs to be investigated.

3.5. CPBF7 bound to β-hexosaminidase, similar to CPBF8, amoebapore and MPR

Three possible lysosomal luminal proteins, two β-hexosaminidases (XP_656208 (EHI_012010) and XP_650273 (EHI_007330)) and an amoebapore B precursor, were detected in the whole immunoprecipitated sample from CPBF7-HA (Table 1), while those were

not detectable by SDS–PAGE and silver staining (Fig. 1). The *E. histolytica* genome encodes three β-hexosaminidases, two of which were bound to CPBF7-HA, and the other, AJ582954 (XP_657529, EHI_148130), was recognised by CPBF8 (Furukawa et al., 2012, Table 1). It was shown that this β-hexosaminidase (AJ582954) is localised in cytoplasmic granules and phagosomes (Riekenberg et al., 2004; Furukawa et al., 2012) and all three β-hexosaminidases have the signal peptide. Thus, unlike amylases, all β-hexosaminidases seem to be carried by CPBFs. Both CPBF7 and CPBF8 have SRR, which was shown to be essential for β-hexosaminidase binding by CPBF8 (Furukawa et al., 2012).

β-Hexosaminidases are involved in the hydrolysis of terminal N-acetyl-D-hexosamine residues in hexosaminides. When *E. histolytica* trophozoites propagate extraintestinally, they take a route similar to that during metastasis of cancer cells (Leroy et al., 1995), which requires both proteases and glycosidases during the passage of the basement membrane (Bernacki et al., 1985; Liotta, 1984). Furthermore, it was shown that β-hexosaminidase activity is involved in mucin degradation (Stewart-Tull et al., 1986). β-Hexosaminidase was found as one of the transcriptionally upregulated genes after *E. histolytica* trophozoite's contact with human colon epithelia in an ex vivo model (Thibeaux et al., 2013). Taken together, β-hexosaminidases and their traffic regulation are important for the pathogenesis of *E. histolytica*.

Identification of amoebapore B precursor as a CPBF7 cargo is important as amoebapores are described as major virulence factors (Leippe et al., 2005). Amoebapores are the cytolytic peptides homologous to granulysin, which is present in human cytotoxic lymphocytes, displays potent cytolytic activity towards bacterial and human cells, and forms ion channels in artificial membranes (Leippe, 1997). Amoebapores are targeted to lysosomes and mainly involved in degradation of ingested bacteria. Inhibition of expression of the *amoebapore A* gene by antisense or gene silencing caused a reduction in virulence, suggesting that this protein plays a key role in pathogenesis (Bracha et al., 1999, 2003).

One of two MPRs in *E. histolytica*, MPR1, was also found as a CPBF7-binding protein. MPR1 is predicted to have a single carbohydrate binding domain (CRD), but the a.a. residues implicated for mannose 6-phosphate binding (Dahms et al., 2008) are not conserved. We performed immunoprecipitation and LC-MS/MS analysis of HA-tagged MPR1 but failed to identify the ligand (Nakada-Tsukui et al., data not shown). It is of note that in *Saccharomyces cerevisiae* Vps10p and a single CRD domain-containing protein, Mrl1p (Whyte and Munro, 2001), cooperatively function in the traffic of lysosomal (vacuole in yeast) proteins, but no ligand was assigned for Mrl1p. It is plausible that MPR1 and CPBF7 are cooperatively involved in trafficking to lysosomes.

3.6. CPBF9 bound to lysozyme, similar to CPBF8

Lysozyme 2, XP_656933 (EHI_096570), was found to bind to CPBF9-HA (Table 1), however it was not detectable by SDS–PAGE or silver staining (Fig. 1). It has previously been shown that lysozyme 2 is also recognised by CPBF8 (Furukawa et al., 2012). Lysozymes are encoded by six independent genes in the *E. histolytica* genome and annotated as lysozymes or N-acetylmuraminidase. Among them, the *lysozyme 2* gene is the most highly transcribed (Penuliar et al., 2012). Lysozymes are well-known glycosidases that degrade the bacterial cell wall (Chipman et al., 1967). It was reported that lysozyme genes were poorly expressed in an avirulent *E. histolytica* Rahman strain and in *Entamoeba dispar* (MacFarlane and Singh, 2006; Davis et al., 2007). Furthermore, expression of lysozyme genes was repressed when *E. histolytica* trophozoites were treated with 5-azacytidine, a potent inhibitor of DNA methyltransferase, and the repression of lysozyme genes correlated with a reduction in virulence (Ali et al., 2008). We also demonstrated that repression of *CPBF8* gene expression by small antisense RNA-mediated transcriptional silencing (Bracha et al., 1999, 2003) caused a decrease in the targeting of lysozyme 2 to phagosomes and delay in digestion of ingested gram-positive bacteria (Furukawa et al., 2012). It was also reported that the SRR of CPBF8 is glycosylated and glycosylation is important for the binding of β-hexosaminidase and lysozyme 2 (Furukawa et al., 2012). CPBF9 has no SRR, and does not seem to have post-translational modifications. These data indicate that mechanisms of interaction between CPBF9 and lysozyme 2 must be different from those of CPBF8 and lysozyme 2.

3.7. Identification of additional CPBF1 binding proteins

To further identify additional lysosomal proteins recognised by CPBF1 other than previously identified CPs, we vigorously searched for other binding proteins. Based on the criteria described in Section 2.6, the proteins identified in four independent experiments and those repeatedly detected (either in two, three or four out of four experiments) are listed (Table 2). We detected a total of 20 proteins in four experiments. Among them, four proteins were detected in all four experiments (EhCP-A2, EhCP-A4, EhCP-A5 and CPBF1 itself), while three other proteins were detected in two or three experiments. EhCP-A1 and EhCP-A6 were detected only in a single experiment (Supplementary Table S3).

None of the possible soluble lysosomal proteins, other than CPs, were detected as CPBF1-HA binding protein, reinforcing the specificity of CPBF1 to CPs and verifying the stringency of the protocol used in the study. We previously identified EhCP-A1 as one of the cargos for CPBF1 by a pull-down experiment of CPBF1-HA, followed by immunoblot analysis using anti-EhCP-A1 antibody (Nakada-Tsukui et al., 2012). One should note that anti-EhCP-A1 antibody cross-reacted with EhCP-A2 due to the high a.a. identity (81%) (Mitra et al., 2007). In the present study, LC–MS/MS data have clearly shown that CPBF1 preferentially interacts with EhCP-A2 but not EhCP-A1. EhCP-A1 and EhCP-A2 are the two major CPs with comparably high expression levels, followed by EhCP-A5 in *E. histolytica* HM-1:IMSS (Tillack et al., 2007). EhCP-A4 is one of the poorly expressed CPs, but suggested to be involved in the pathogenesis of invasive amebiasis (Tillack et al., 2007; He et al., 2010). Reproducible detection of EhCP-A4 in all of the experiments indicates the high affinity of CPBF1 toward EhCP-A4. Interestingly, EhCP-A4 is localised to the nuclear region and the acidic compartment (He et al., 2010). The role of CPBF1 in the EhCP-A4 localization needs to be elucidated. As more than 95% of the CP activity of *E. histolytica* trophozoites is attributed to EhCP-A1, A2, A5 and A7 (Bruchhaus et al., 2003; Irmer et al., 2009), the amounts of CPs bound to CPBF1 does not seem to be proportional to their

Table 2
Reproducibility of identified cysteine protease binding protein family 1 (CPBF1) binding proteins.

Identified proteins	MW	Accession number		Number of experiments in which the protein was identified
		GenBank	AmoebaDB	
CPBF1	101 kDa	XP_655218	EHI_164800	4
EhCP-A2	35 kDa	XP_650642	EHI_033710	4
EhCP-A4	34 kDa	XP_656602	EHI_050570	4
EhCP-A5	35 kDa	XP_650937	EHI_168240	4
70 kDa heat shock protein	73 kDa	XP_654737	EHI_199590	3
Mannosyltransferase	49 kDa	XP_650080	EHI_029580	2
Hypothetical protein	43 kDa	XP_649888	EHI_146110	2

EhCP, *Entamoeba histolytica* cysteine protease.

expression levels, but determined by the ligand specificity of CPBF1.

We reproducibly identified heat shock protein 70 (Hsp70) (XP_654737, EHI_199590), which has the ER retention signal (KDEL) at the carboxyl terminus (three out of four experiments). It is worth noting that this protein was repeatedly detected in all immunoprecipitation experiments except for CPBF4. Mannosyltransferase, localised in the ER (Maeda and Kinoshita, 2008; Loibl and Strahl, 2013), was also repeatedly identified (two out of four experiments). Identification of the ER-residing Hsp70 and mannosyltransferase suggests possible involvement of ER proteins in the functionality of CPBF1. Hypothetical protein (XP_649888, EHI_146110), with no detectable domain or motif, was detected in two out of four experiments.

3.8. Analysis of ligands for CPBF3, CPBF4, CPBF5 and CPBF11

No known or possible hydrolases or membrane proteins were detected either by SDS–PAGE analysis followed by silver staining or LC–MS/MS analysis of the whole immunoprecipitated samples, with a few exceptions: EhCP-A2 and a light subunit of galactose/N-acetylgalactosamine-inhibitable lectin in CPBF4-HA (with low QV, 3.44) (Fig. 1, Table 1). Thus, no specific ligand was identified for CPBF4. It may be worth noting that the pIs of CPBF3, CPBF4, CPBF11 (7.2, 6.5 and 6.5, respectively) are higher than those of other members; the average pI value of the 11 CPBFs is 5.5.

CPBF3 was detected by immunoprecipitation of CPBF4-HA and vice versa. CPBF3 and CPBF4 have high mutual a.a. identity (75%, Supplementary Table S4). Three peptides detected in the CPBF3-HA pull-down sample were mapped to CPBF4 (5% coverage). Similarly, five peptides were mapped to CPBF3 in the CPBF4-HA pull-down sample (7% coverage). These data indicate interaction between CPBF3 and CPBF4.

Serine threonine isoleucine rich protein (STIRP) was found in the immunoprecipitated sample from CPBF4-HA. Another isotype of EhSTIRP (XP_656227.2, EHI_012330) was also detected, although it was removed from Table 1 due to lack of the signal peptide. Since the carboxyl-terminal regions of these proteins show high mutual similarity (MacFarlane and Singh, 2007), detected peptides did not differentiate two EhSTIRPs. EhSTIRP, which contains a single transmembrane domain, was exclusively expressed in virulent *E. histolytica* strains, but not in non-virulent *E. histolytica* Rahman strain or *E. dispar*, and thus is considered to be a virulent-associated protein (MacFarlane and Singh, 2007). Possible interaction between EhSTIRP and CPBF4 needs to be further verified.

A light subunit of galactose/N-acetylgalactosamine-inhibitable lectin was found in the immunoprecipitated sample from CPBF4-HA and CPBF5-HA. It is well established that this lectin is involved in the interaction between *E. histolytica* and host cells/microbes, and is essential for pathogenesis (Ravdin et al., 1989; Petri et al., 2002). The lectin is composed of three subunits, i.e. heavy, intermediate and light subunits (Petri et al., 2002). The 170 kDa heavy subunit with a transmembrane domain and the 31–35 kDa glycosylphosphatidylinositol (GPI)-anchored light subunit form a heterodimer by disulfide bonds. An intermediate subunit of 150 kDa is non-covalently associated with the heterodimer. All three subunits are encoded by multigene families. There are five genes for the heavy subunit, six to seven for the light subunit and 30 for the intermediate subunit (Petri et al., 2002). The fact that only specific light subunits were associated with CPBF4 and CPBF5, respectively, indicates that these light subunits together with the corresponding CPBFs may be involved in trafficking of the surface receptor in association with other lysosomal receptors.

CPBF5 was found to also interact with two additional proteins, neither of which seems to be a potential lysosomal protein. Interestingly, an immunofluorescence assay (Fig. 2, see Section 3.9) showed that CPBF5-HA is localised in lysosomes, as indicated by colocalization with Lysotracker. This is in good contrast with other CPBFs mainly localised in the ER/Golgi compartments, e.g., CPBF1, CPBF6 and CPBF8. If it is assumed that the ligand-receptor binding is affected by pH, as shown for CPBF1 (Nakada-Tsukui et al., 2012), the conditions for pull-down experiments may need to be further optimised to obtain the ligand of CPBF5.

A 45 kDa band was specifically detected in the immunoprecipitated sample from CPBF11-HA by SYPRO ruby staining (Fig. 1), but identified as CPFB11 itself by LC-MS/MS analysis (Supplementary Table S5). The whole immunoprecipitated sample was subjected to MS analysis, but no additional binding protein was detected (Table 1). It is worth noting that mRNA expression of a gene encoding a CPBF11 homologue in *Entamoeba invadens*, a reptilian sibling of *E. histolytica*, is 4.7–9-fold upregulated after 24–120 h of encystation (De Cádiz et al., 2013). The finding may explain why no CPBF11 ligand was discovered in trophozoites. Identification of CPBF11 binding proteins from *E. histolytica* cysts may be needed.

3.9. Intracellular localization of CPBFs

Intracellular localization of CPBFs was examined by immnofluorescence assay with anti-HA antibody using Lysotracker red stained trophozoites of CPBF-HA-expressing lines (Fig. 2). CPBF2, 7, 9 and 10-HA were detected on vacuolar membranes and small membrane structures scattered all over the cells. CPBF3, 4 and 11 were mostly localised on small membrane structures and hardly detected on vacuolar membranes. In contrast, as briefly mentioned 3.8, CPBF5-HA was nicely colocalized with Lysotracker red, indicating lysosomal localization. However, CPBF5 was not identified in our previous phagosome proteome study (Okada et al., 2006; Furukawa et al., 2012), which may be due to low expression of endogenous CPBF5. Partial colocalization was also observed for CPBF2, 7, 9 and 10.

Localization of the CPBF proteins involved in the transport of carbohydrate digesting enzymes, CPBF2, CPBF7 and CPBF10, was similar to that of CPBF6 and CPBF8 (Furukawa et al., 2012, 2013). They are localised on both the vacuolar membrane and the small membrane structures. It was previously shown that CPBF6 and CPBF8 are co-localised with pyridine nucleotide transhydrogenase (PNT), which utilises the electrochemical proton gradient across the membrane to drive NADPH formation from NADH (Yousuf et al., 2010).

3.10. Structure of CPBFs

To infer structural/functional domains in CPBFs we used FORTE (Tomii and Akiyama, 2004), which performs profile-profile alignments for protein structure prediction. FORTE allowed us to identify five PPC (bacterial prepeptidase carboxyl-terminal domain)-like domains at the luminal portion of each CPBF. We show here, as an example, the alignment of the putative amino-terminal PPC domain (D1) of CPBF1 and the bacterial collagen-binding domain (PDB code: 1NQJ) (Wilson et al., 2003) with the highest Z-score calculated by FORTE (Fig. 3A). 1NQJ belongs to the PPC family (Pfam ID: PF04151) (Punta et al., 2012), which includes a large and diverse set of protein domains that possess two β-sheets. The PPC domains are typically located at the carboxyl-termini of secreted proteases and may be involved in their secretion and/or localization (Yeats et al., 2003). By manual inspection and sequence alignment, we identified an additional PPC-like domain, D4, which was not inferred by FORTE. There are similarities between individual PPC-like domains of each CPBF. We identified conserved cysteines and aromatic/hydrophobic residues in the predicted β-strands of the six PPC-like domains of CPBF1 (Supplementary Fig. S2). Similarly, it appears that all CPBFs contain a region consisting of six PPC-like domains in the luminal portion.

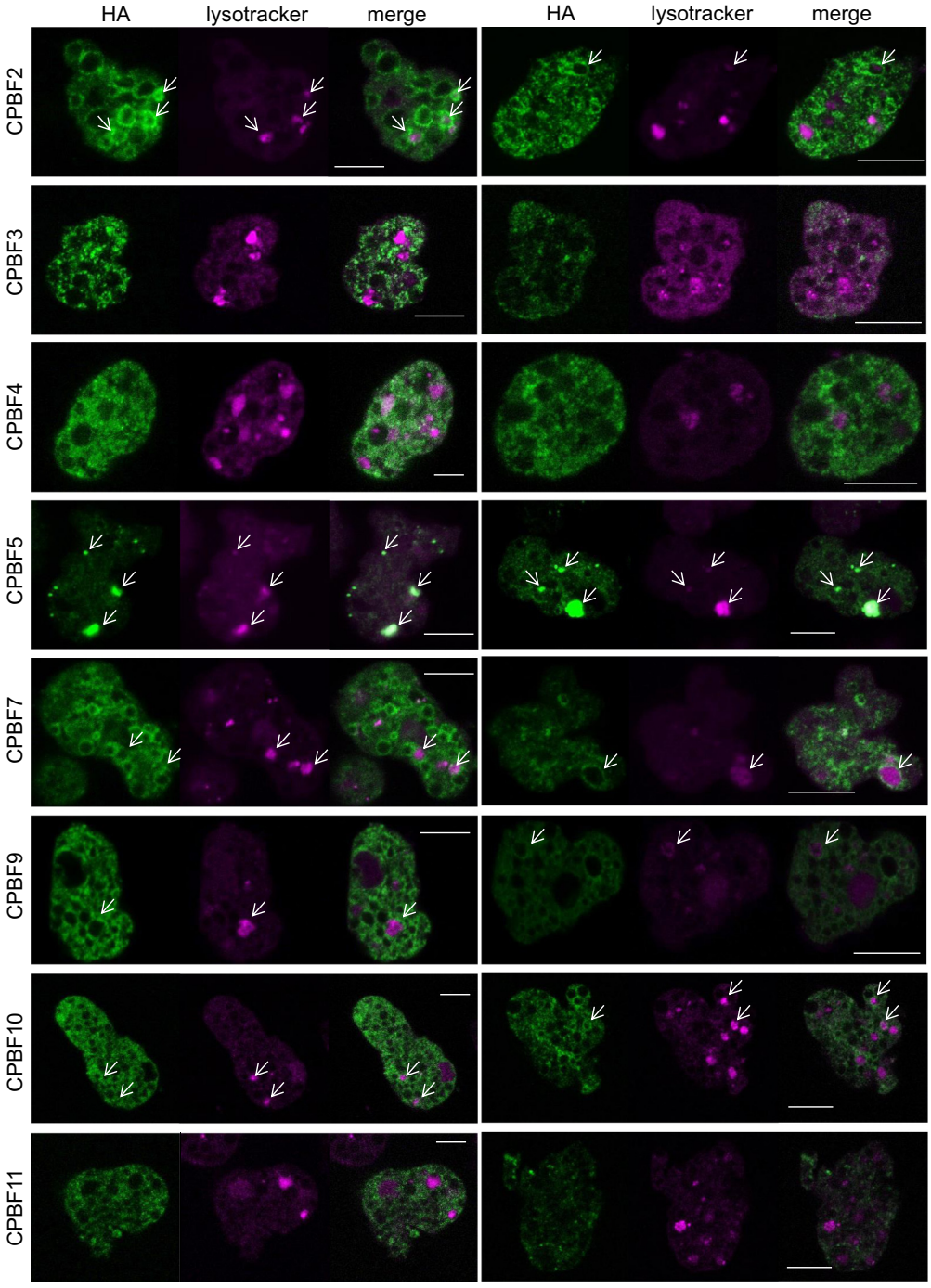

Fig. 2. Immunofluorescence images of *Entamoeba histolytica* cysteine protease binding protein families (CPBFs) 2, 3, 4, 5, 7, 9, 10 and 11. Trophozoites of the CPBF-haemagglutinin (HA)-expressing transformants were incubated with Lysotracker Red, fixed, reacted with anti-HA antibody and confocal images were captured on LSM510. Thirteen to 61 cells were examined in one to five independent experiments for each CPBF. Two representative cells are shown for each CPBF. Arrows depict Lysotracker accumulation in the CPBF-positive vesicles and vacuoles. Bars = 10 μm.

A phylogenetic analysis of six PPC-like domains of 11 CPBFs indicates that corresponding domains (e.g., D3) of all CPBFs tend to form clusters (with limited bootstrap support) (Fig. 3B). This likely implies, together with the fact that all CPBFs have similar domain configuration, that individual corresponding domains (e.g., D3, D5) retain distinct structural role(s).

3.11. PPC domain is a functional unit of the ligand binding of CPBF1

To investigate whether the binding activity of CPBF1 to CP can be attributable to specific PPC domain(s), each PPC domain, CPBF1 domains 1–6 (D1–D6), was expressed as GST-fusion protein in *E. coli*, with CPBF8 domain 1 (D1) as negative control and an in vitro pull-down assay was performed (Fig. 4). Among the six PPC domains of CPBF1, D3 showed significantly higher affinity ($P < 0.05$) compared with D1 and D6 (Fig. 4C). D5 also showed significantly higher affinity than D6 ($P < 0.05$). These results indicate that single PPC domains per se have the ability of ligand binding. D4 was truncated or degraded during expression and/or purification and not used in the study (data not shown).

The mechanisms of ligand recognition of CPBFs have not been elucidated. We previously showed that carbohydrate modifications

A

```
Conf        9860476113478336888627777765405873143364314653446O
Pred        CCCCEEEECCCCCCEEEEEECCCCCCCCCCEEEECCCCCCCCCCCCCCE
C1D1   21 CQTAHEISLADLDENKVASFDGDTTKQTNFCWVCNPRFGCGVDENLFPRS   70
1NQJ  908 --KATVIPNFNTTMQGSLLGD-------------------------DSR  929
DSSP        EE       EEEEE                              E

Conf        6999977999759998236656335158996110478532222356776S
Pred        EEEEEECCCCCEEEEEECCCCCCCCCEEEEEEECCCCCCCCCCCCCCCCC
C1D1   71 AFYKFTLKQNMPIEISTCNTYTNYNTRIAVLTTCEANRAHTCASSNDDDG  120
1NQJ  930 DYYSFEVKEEGEVNIELDKKDEF-GVTWTLHP-----ES---DRITYGQV  973
DSSP        EEEEEEE   EEEEEEEEE     EEEEEEEE            EE EE

Conf        6666676202689982378769999972654434799999819
Pred        CCCCCCCCCEEEEEEECCCCEEEEEEECCCCCEEEEEEEEEC
C1D1  121 EISVCAGGKSQLVFNAVAGTTYYIVVSGTTASDVGVFRVTVKK  163
1NQJ  974 DG------NKVSNKVKLRPGKYYLLVYKYS--GSGNYELRVNK 1008
DSSP        E        EEEEEEEE  EEEEEEEEEEE    EEEEEEEE
```

B

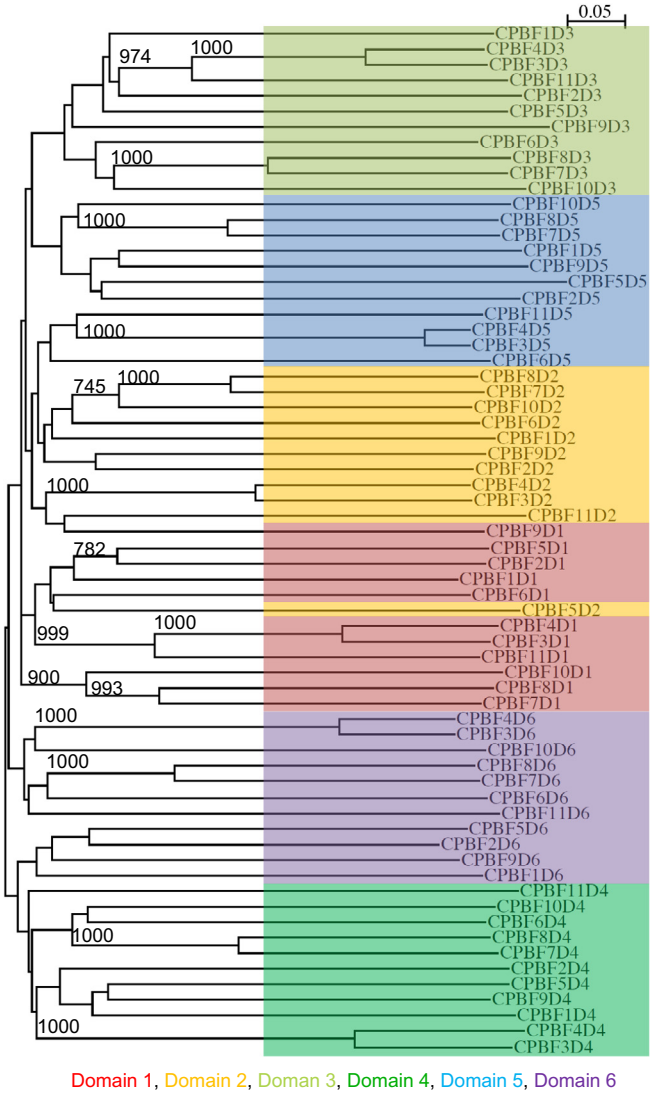

Fig. 3. Prediction of the prepeptidase carboxyl-terminal domain in *Entamoeba histolytica* cysteine protease binding protein families (CPBFs) by FORTE. (A) Amino acid sequence alignment of the putative N-terminal prepeptidase carboxyl-terminal domain of CPBF1 (shown as C1D1) and a bacterial collagen-binding domain (PDB code: 1NQJ), constructed by FORTE. Predicted secondary structure and its confidence value, at each residue of the prepeptidase carboxyl domain, calculated by PSIPRED (McGuffin et al., 2000), are indicated in the "Pred" row and the "Conf" row, respectively. A greater value means higher confidence. In the "Pred" row, "E" indicates a β-sheet residue. In the "DSSP" row, the secondary structure assignments for the structure of 1NQJ, determined by database of secondary structure assignments (DSSP), are shown. Predicted and actual β-sheet residues are coloured in orange. (B) Phylogenetic analysis of six prepeptidase carboxyl domains from CPBFs. Confidence (bootstrap) values (only ⩾ 700), at each branch, from 1000 resamplings are shown.

of SRR are involved in ligand binding of CPBF6 and CPBF8 (Furukawa et al., 2012, 2013). However, only CPBF6-8 apparently have SRR, whereas other CPBFs lack it. Further structural studies are required to better understand the mechanisms of ligand recognition, binding and dissociation, as well as ligand specificities.

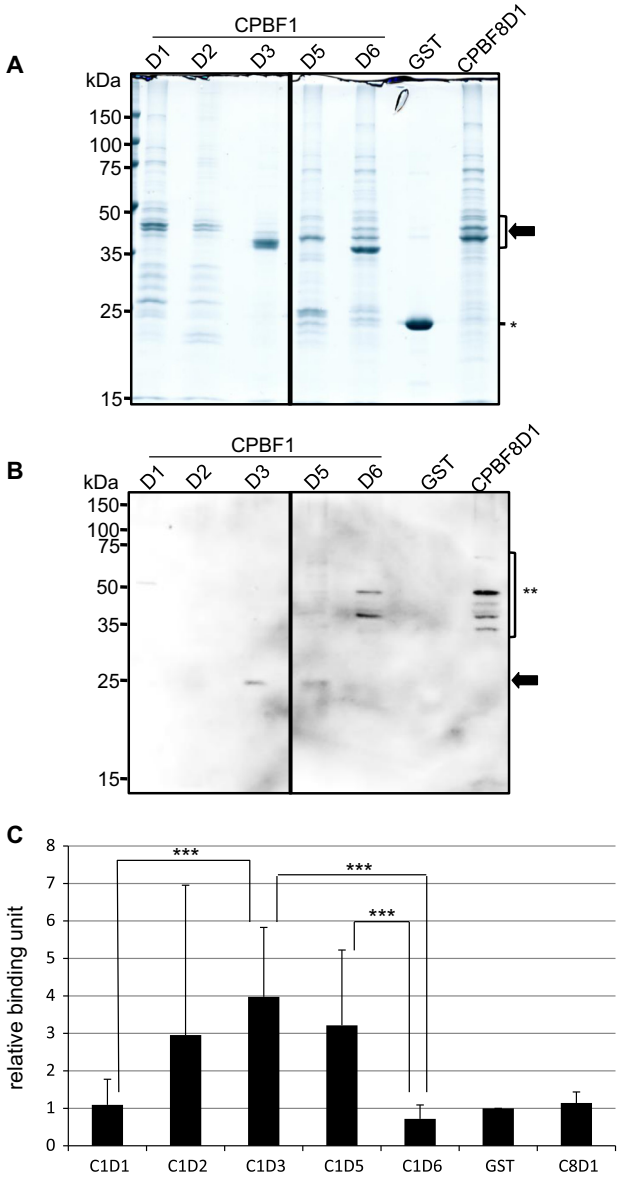

Fig. 4. Binding assay of individual domains of *Entamoeba histolytica* cysteine protease binding protein family (CPBF) 1 to cysteine protease (CP)-A5. GST-fused recombinant proteins containing each prepeptidase carboxyl-terminal (PPC) domain (D1, 2, 3, 5 and 6) from CPBF1 were mixed with *E. histolytica* lysates and purified with glutathione-conjugated beads. The CPBF/ligand mixtures were separated by SDS–PAGE and either stained by Coomassie Brilliant Blue staining or subjected to immunoblot analysis using anti-CP-A5 antibody. Note that GST-only and GST fused with D1 from CPBF8 were used as negative controls. D4 was not used in this assay because a large proportion of GST-CPBF1 D4 was degraded during production or purification. Note that images shown in A and B were cropped from a single image and combined. (A) Coomassie Brilliant Blue staining. An arrow indicates GST-fused CPBF1 PPC domain recombinant (CPBF1 D1-D6) and CPBF8 D1 (an irrelevant control) used for pull down assays. *GST control. (B) Immunoblot analysis. An arrow indicates CP-A5. **Non-specific bands. (C) Quantification of relative binding efficiency of individual prepeptidase carboxyl domains to CP-A5. Relative binding efficiency of each GST-prepeptidase carboxyl domain fusion protein to CP-A5 was expressed after normalisation against the value of the GST control (set to 1). S.D.s of three replicates are shown with error bars. ***Statistical significance ($P < 0.05$ by Student's t test).

Acknowledgements

We thank Ms Courtney Spears and Dr. Andrew J. Roger, Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biochemistry and Molecular Biology at Dalhousie University, Canada for *Mastigamoeba* genome information. We thank Dr. Nicholas E. Sherman, W.M. Keck Biomedical Mass Spectrometry Laboratory, University of Virginia, USA for mass spectrometric analysis. This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan to T.N. (23117001, 23117005, 23390099) and K.N.-T. (24590513), a Grant-in-Aid on Bilateral Programs of Joint Research Projects and Seminars from Japan Society for the Promotion of Science, a Grant-in-Aid on Strategic International Research Cooperative Program from Japan Science and Technology Agency, a grant for research on emerging and re-emerging infectious diseases from the Ministry of Health, Labour and Welfare of Japan (H23-Shinko-ippan-014) to T.N., a grant for research to promote the development of anti-AIDS pharmaceuticals from the Japan Health Sciences Foundation (KHA1101) to T.N., Strategic International Research Cooperative Program from Japan Science and Technology Agency to T.N. and by Global COE Program (Global COE for Human Metabolomic Systems Biology) from MEXT, Japan to T.N., Platform for Drug Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture, Sports, Science and Technology, Japan to KT and KN-T.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijpara.2014.04.008.

References

Ackers, J.P., Mirelman, D., 2006. Progress in research on *Entamoeba histolytica* pathogenesis. Curr. Opin. Microbiol. 9, 367–373.

Ali, I.K.M., Ehrenkaufer, G.M., Hackney, J.A., Singh, U., 2008. Growth of the protozoan parasite Entamoeba histolytica in 5-azacytidine has limited effects on parasite gene expression. BMC Genomics 8, 7.

Bernacki, R.J., Niedbala, M.J., Korytnyk, W., 1985. Glycosidases in cancer and invasion. Cancer Metastasis Rev. 4, 81–101.

Bracha, R., Nuchamowitz, Y., Leippe, M., Mirelman, D., 1999. Antisense inhibition of amoebapore expression in *Entamoeba histolytica* causes a decrease in amoebic virulence. Mol. Microbiol. 34, 463–472.

Bracha, R., Nuchamowitz, Y., Mirelman, D., 2003. Transcriptional silencing of an amoebapore gene in *Entamoeba histolytica*: molecular analysis and effect on pathogenicity. Eukaryot. Cell 2, 295–305.

Brinen, L.S., Que, X., McKerrow, J.H., Reed, S.L., 2000. Homology modeling of *Entamoeba histolytica* cysteine proteinases reveals the basis for cathepsin L-like structure with cathepsin B-like specificity. Arch. Med. Res. 31, S63–S64.

Bruchhaus, I., Loftus, B.J., Hall, N., Tannich, E., 2003. The intestinal protozoan parasite *Entamoeba histolytica* contains 20 cysteine protease genes, of which only a small subset is expressed during in vitro cultivation. Eukaryot. Cell 2, 501–509.

Chipman, D.M., Grisaro, V., Sharon, N., 1967. The binding of oligosaccharides containing N-acetylglucosamine and N-acetylmuramic acid to lysozyme. J. Biol. Chem. 242, 4388–4394.

Clark, C.G., Diamond, L.S., 2002. Methods for cultivation of luminal parasitic protists of clinical importance. Clin. Microbiol. Rev. 15, 329–341.

Dahms, N.M., Olson, L.J., Kim, J.J., 2008. Strategies for carbohydrate recognition by the mannose 6-phosphate receptors. Glycobiology 18, 664–678.

Davis, P.H., Scholze, J., Stanley Jr., S.L., 2007. Transcriptomic comparison of two *Entamoeba histolytica* strains with defined virulence phenotypes identifies new virulence factor candidates and key differences in the expression patterns of cysteine proteases, lectin light chains, and calmodulin. Mol. Biochem. Parasitol. 151, 118–128.

De Cádiz, A.E., Jeelani, G., Nakada-Tsuku, i.K., Caler, E., Nozaki, T., 2013. Transcriptome analysis of encystation in *Entamoeba invadens*. PLoS One 8, e74840.

Diamond, L.S., Harlow, D.R., Cunnick, C.C., 1978. A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. Trans. R. Soc. Trop. Med. Hyg. 72, 431–432.

Furukawa, A., Nakada-Tsukui, K., Nozaki, T., 2012. Novel transmembrane receptor involved in phagosome transport of lysozymes and β-hexosaminidase in the enteric protozoan *Entamoeba histolytica*. PLoS Pathog. 8, e1002539.

Furukawa, A., Nakada-Tsukui, K., Nozaki, T., 2013. Cysteine protease-binding protein family 6 mediates the trafficking of amylases to phagosomes in the enteric protozoan *Entamoeba histolytica*. Infect. Immun. 81, 1820–1829.

Gilchrist, C.A., Houpt, E., Trapaidze, N., Fei, Z., Crasta, O., Asgharpour, A., Evans, C., Martino-Catt, S., Baba, D.J., Stroup, S., Hamano, S., Ehrenkaufer, G., Okada, M., Singh, U., Nozaki, T., Mann, B.J., Petri Jr., W.A., 2006. Impact of intestinal colonization and invasion on the *Entamoeba histolytica* transcriptome. Mol. Biochem. Parasitol. 147, 163–176.

He, C., Nora, G.P., Schneider, E.L., Kerr, I.D., Hansell, E., Hirata, K., Gonzalez, D., Sajidm, M., Boyd, S.E., Hruz, P., Cobo, E.R., Le, C., Liu, W.T., Eckmann, L., Dorrestein, P.C., Houpt, E.R., Brinen, L.S., Craik, C.S., Roush, W.R., McKerrow, J., Reed, S.L., 2010. A novel *Entamoeba histolytica* cysteine proteinase, EhCP4, is key for invasive amebiasis and a therapeutic target. J. Biol. Chem. 285, 18516–18527.

Hellberg, A., Nickel, R., Lotter, H., Tannich, E., Bruchhaus, I., 2001. Overexpression of cysteine proteinase 2 in *Entamoeba histolytica* or *Entamoeba dispar* increases amoeba-induced monolayer destruction in vitro but does not augment amoebic liver abscess formation in gerbils. Cell. Microbiol. 3, 13–20.

Hellberg, A., Nowak, N., Leippe, M., Tannich, E., Bruchhaus, I., 2002. Recombinant expression and purification of an enzymatically active cysteine proteinase of the protozoan parasite *Entamoeba histolytica*. Protein Expr. Purif. 24, 131–137.

Irmer, H., Tillack, M., Biller, L., Handal, G., Leippe, M., Roeder, T., Tannich, E., Bruchhaus, I., 2009. Major cysteine peptidases of *Entamoeba histolytica* are required for aggregation and digestion of erythrocytes but are dispensable for phagocytosis and cytopathogenicity. Mol. Microbiol. 72, 658–667.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

Leippe, M., 1997. Amoebapores. Parasitol. Today 13, 178–183.

Leippe, M., Bruhn, H., Hecht, O., Grötzinger, J., 2005. Ancient weapons: the three-dimensional structure of amoebapore A. Trends Parasitol. 21, 5–7.

Leroy, A., Mareel, M., De Bruyne, G., Bailey, G., Nelis, H., 1995. Metastasis of *Entamoeba histolytica* compared to colon cancer: one more step in invasion. Inv. Metastasis 14, 177–191.

Liotta, L.A., 1984. Tumor invasion and metastases: role of the basement membrane. Am. J. Pathol. 117, 339–348.

Loibl, M., Strahl, S., 2013. Protein O-mannosylation: what we have learned from baker's yeast. Biochim. Biophys. Acta 1833, 2438–2446.

MacFarlane, R.C., Singh, U., 2006. Identification of differentially expressed genes in virulent and nonvirulent *Entamoeba* species: potential implications for amebic pathogenesis. Infect. Immun. 74, 340–351.

MacFarlane, R.C., Singh, U., 2007. Identification of an *Entamoeba histolytica* serine-, threonine-, and isoleucine-rich protein with roles in adhesion and cytotoxicity. Eukaryot. Cell 6, 2139–2146.

Maeda, Y., Kinoshita, T., 2008. Dolichol-phosphoate mannose synthase: structure, function and regulation. Biochim Biophys Acta. 1780, 861–868.

McGuffin, L.J., Bryson, K., Jones, D.T., 2000. The PSIPRED protein structure prediction server. Bioinformatics 16, 404–405.

Meléndez-López, S.G., Herdman, S., Hirata, K., Choi, M.H., Choe, Y., Craik, C., Caffrey, C.R., Hansell, E., Chávez-Munguía, B., Chen, Y.T., Roush, W.R., McKerrow, J., Eckmann, L., Guo, J., Stanley Jr, S.L., Reed, S.L., 2007. Use of recombinant *Entamoeba histolytica* cysteine proteinase 1 to identify a potent inhibitor of amebic invasion in a human colonic model. Eukaryot. Cell 6, 1130–1136.

Mitra, B.N., Saito-Nakano, Y., Nakada-Tsukui, K., Sato, D., Nozaki, T., 2007. Rab11B small GTPase regulates secretion of cysteine proteases in the enteric protozoan parasite *Entamoeba histolytica*. Cell. Microbiol. 9, 2112–2125.

Nakada-Tsukui, K., Okada, H., Mitra, B.N., Nozaki, T., 2009. Phosphatidylinositol-phosphates mediate cytoskeletal reorganization during phagocytosis via a unique modular protein consisting of RhoGEF/DH and FYVE domains in the parasitic protozoon *Entamoeba histolytica*. Cell. Microbiol. 11, 1471–1491.

Nakada-Tsukui, K., Saito-Nakano, Y., Ali, V., Nozaki, T., 2005. A retromerlike complex is a novel Rab7 effector that is involved in the transport of the virulence factor cysteine protease in the enteric protozoan parasite *Entamoeba histolytica*. Mol. Biol. Cell 16, 5294–5303.

Nakada-Tsukui, K., Tsuboi, K., Furukawa, A., Yamada, Y., Nozaki, T., 2012. A novel class of cysteine protease receptors that mediate lysosomal transport. Cell. Microbiol. 14, 1299–1317.

Nakatsu, F., Ohno, H., 2003. Adaptor protein complexes as the key regulators of protein sorting in the post-Golgi network. Cell Struct. Funct. 28, 419–429.

Nozaki, T., Asai, T., Sanchez, L.B., Kobayashi, S., Nakazawa, M., Takeuchi, T., 1999. Characterization of the gene encoding serine acetyltransferase, a regulated enzyme of cysteine biosynthesis from the protist parasites *Entamoeba histolytica* and *Entamoeba dispar*. Regulation and possible function of the cysteine biosynthetic pathway in *Entamoeba*. J. Biol. Chem. 274, 32445–32452.

Okada, M., Huston, C.D., Oue, M., Mann, B.J., Petri Jr., W.A., Kita, K., Nozaki, T., 2006. Kinetics and strain variation of phagosome proteins of *Entamoeba histolytica* by proteomic analysis. Mol. Biochem. Parasitol. 145, 171–183.

Penuliar, G.M., Furukawa, A., Nakada-Tsukui, K., Husain, A., Sato, D., Nozaki, T., 2012. Transcriptional and functional analysis of trifluoromethionine resistance in *Entamoeba histolytica*. J. Antimicrob. Chemother. 67, 375–386.

Perrière, G., Gouy, M., 1996. WWW-Query: an on-line retrieval system for biological sequence banks. Biochimie 78, 364–369.

Petri Jr., W.A., Haque, R., Mann, B.J., 2002. The bittersweet interface of parasite and host: lectin-carbohydrate interactions during human invasion by the parasite *Entamoeba histolytica*. Annu. Rev. Microbiol. 56, 39–64.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn R.D., 2012. The Pfam protein families database. Nucl. Acids Res. 40 (Database issue), D290–D301.

Que, X., Kim, S.H., Sajid, M., Eckmann, L., Dinarello, C.A., McKerrow, J.H., Reed, S.L., 2003. A surface amebic cysteine proteinase inactivates interleukin-18. Infect. Immun. 71, 1274–1280.

Que, X., Reed, S.L., 2000. Cysteine proteinases and the pathogenesis of amebiasis. Clin. Microbiol. Rev. 13, 196–206.

Ralston, K.S., Petri Jr., W.A., 2011. Tissue destruction and invasion by *Entamoeba histolytica*. Trends Parasitol. 27, 254–263.

Ravdin, J.I., Stanley, P., Murphy, C.F., Petr, i W.A. Jr.,1989. Characterization of cell surface carbohydrate receptors for *Entamoeba histolytica* adherence lectin. Infect. Immun. 57, 2179–2186.

Riekenberg, S., Flockenhaus, B., Vahrmann, A., Müller, M.C., Leippe, M., Kiess, M., Scholze, H., 2004. The beta-N-acetylhexosaminidase of *Entamoeba histolytica* is composed of two homologous chains and has been localized to cytoplasmic granules. Mol. Biochem. Parasitol. 138, 217–225.

Sambrook, J., Russell, D.W., 2001. Molecular Cloning. Cold Spring Harbor Laboratory Press, New York.

Shiozawa, K., Maita, N., Tomii, K., Seto, A., Goda, N., Akiyama, Y., Shimizu, T., Shirakawa, M., Hiroaki, H., 2004. Structure of the N-terminal domain of PEX1 AAA-ATPase. Characterization of a putative adaptor binding domain. J. Biol. Chem. 279, 50060–50068.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2004. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.

Stewart-Tull, D.E., Ollar, R.A., Scobie, T.S., 1986. Studies on the *Vibrio cholerae* mucinase complex. I. Enzymic activities associated with the complex. J. Med. Microbiol. 22, 325–333.

Thibeaux, R., Weber, C., Hon, C.C., Dillies, M.A., Avé, P., Coppée, J.Y., Labruyère, E., Guillén, N., 2013. Identification of the Virulence Landscape Essential for *Entamoeba histolytica* Invasion of the Human Colon. PLoS Pathog. 9, e1003824.

Tillack, M., Biller, L., Irmer, H., Freitas, M., Gomes, M.A., Tannich, E., Bruchhaus, I., 2007. The *Entamoeba histolytica* genome: primary structure and expression of proteolytic enzymes. BMC Genom. 8, 170.

Tomii, K., Akiyama, Y., 2004. FORTE: a profile–profile comparison tool for protein fold recognition. Bioinformatics 20, 594–595.

Tomii, K., Hirokawa, T., Motono, C., 2005. Protein structure prediction using a variety of profile libraries and 3D verification. Proteins 61 (Suppl. 7), 114–121.

Tomii, K., Sawada, Y., Honda, S., 2012. Convergent evolution in structural elements of proteins investigated using cross profile analysis. BMC Bioinformatics 13, 11.

Wang, G., Jin, Y., Dunbrack Jr., R.L., 2005. Assessment of fold recognition predictions in CASP6. Proteins 61 (Suppl. 7), 46–66.

Whyte, J.R., Munro, S., 2001. A yeast homolog of the mammalian mannose 6-phosphate receptors contributes to the sorting of vacuolar hydrolases. Curr. Biol. 11, 1074–1078.

Wilson, J.J., Matsushita, O., Okabe, A., Sakon, J., 2003. A bacterial collagen-binding domain with novel calcium-binding motif controls domain orientation. EMBO J. 22, 1743–1752.

Yeats, C., Bentley, S., Bateman, A., 2003. New knowledge from old: in silico discovery of novel protein domains in *Streptomyces coelicolor*. BMC Microbiol. 3, 3.

Yousuf, M.A., Mi-ichi, F., Nakada-Tsukui, K., Nozaki, T., 2010. Localization and targeting of an unusual pyridine nucleotide transhydrogenase in *Entamoeba histolytica*. Eukaryot. Cell 9, 926–933.